



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/884,403	06/19/2001	Shamim A. Alpha	27252.2	2286
7590 01/30/2006			EXAMINER	
Petar Kraguljac McDonald Hopkins Co LPA 600 Superior Avenue E Suite 2100 Cleveland, OH 44114			SPOONER, LAMONT M	
			ART UNIT	PAPER NUMBER
			2654	

DATE MAILED: 01/30/2006

Please find below and/or attached an Office communication concerning this application or proceeding.

<b>Office Action Summary</b>	<b>Application No.</b> 09/884,403	<b>Applicant(s)</b> ALPHA, SHAMIM A.	
	<b>Examiner</b> Lamont M. Spooner	<b>Art Unit</b> 2654	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

#### Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

#### Status

- 1) ☒ Responsive to communication(s) filed on 03 November 2005.
- 2a) ☒ This action is **FINAL**.                      2b) ☐ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

#### Disposition of Claims

- 4) ☒ Claim(s) 1 and 4-21 is/are pending in the application.
- 4a) Of the above claim(s) \_\_\_\_\_ is/are withdrawn from consideration.
- 5) ☐ Claim(s) \_\_\_\_\_ is/are allowed.
- 6) ☒ Claim(s) 1 and 4-21 is/are rejected.
- 7) ☐ Claim(s) \_\_\_\_\_ is/are objected to.
- 8) ☐ Claim(s) \_\_\_\_\_ are subject to restriction and/or election requirement.

#### Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☒ The drawing(s) filed on 19 June 2001 is/are: a) ☒ accepted or b) ☐ objected to by the Examiner.  
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).  
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

#### Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All    b) ☐ Some \*    c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
  2. ☐ Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.
  3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).
- \* See the attached detailed Office action for a list of the certified copies not received.

#### Attachment(s)

- |  |   |
|--|---|
| 1) <input type="checkbox"/> Notice of References Cited (PTO-892)   | 4) <input type="checkbox"/> Interview Summary (PTO-413)<br>Paper No(s)/Mail Date. _____ |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948)                                   | 5) <input type="checkbox"/> Notice of Informal Patent Application (PTO-152)             |
| 3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)<br>Paper No(s)/Mail Date _____ | 6) <input type="checkbox"/> Other: _____  |

## **DETAILED ACTION**

### ***Response to Arguments***

1. In response to applicant's arguments, see remarks 11/3/05, the objection to claim 10 is withdrawn.
2. Applicant's arguments filed 11/3/05, regarding claim 15 have been fully considered but they are not persuasive.

In response to applicant's arguments, p.11.para4, "Pon does not teach or suggest disproving a probability ..." The Examiner cannot concur. It is inherent to a positive step of proving a probability assumption, that disproving a probability assumption is also realized, more specifically, C.6.line 65-C.7.line 22-the "statistic that indicates whether a selected word is in a chosen language", wherein the "probability that a character string belongs to each of the candidate languages result inherently determines the value that a character string does not belong. Probability values range from 0 to 1, by definition of probability, therefore the Examiner cannot concur with applicant's conclusion that, (1 or 0) are not probabilities, "Pon does not use probabilities...". Accordingly, applicant's arguments pertaining to claims 16-20 are also not persuasive.

3. Applicant's arguments with respect to claims 1-14 have been considered but are moot in view of the new ground(s) of rejection.

### ***Claim Rejections - 35 USC § 102***

4. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

Art Unit: 2654

A person shall be entitled to a patent unless –

(b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

(e) the invention was described in (1) an application for patent, published under section 122(b), by another filed in the United States before the invention by the applicant for patent or (2) a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a) shall have the effects for purposes of this subsection of an application filed in the United States only if the international application designated the United States and was published under Article 21(2) of such treaty in the English language.

5. Claims 1, 4-13 are rejected under 35 U.S.C. 102(e) as being anticipated by Elworthy (US 6,125,362).

As per **claim 1**, Elworthy teaches a system for automatically determining a language of a document from a set of candidate languages, the system comprising:

a database containing probability data for a plurality of text strings each having a predetermined length equal to each other (C.5.lines 6, 7, 11, 12- his individual letters, and as bi-grams, as predetermined length of 2, elements/group, C.7.lines 33-35-as his database), each text string of the plurality of text strings having an associated probability value indicating a probability that the text string occurs within a language based on occurrences of the text string in all of the candidate languages (C.7.lines 50-65)

logic for setting a negative assumption value for each of the candidate languages indicating the document is not one of the candidate languages (C.7.lines 60-65- wherein the “probability that a character string belongs to each of the candidate languages result inherently determines the value that a character string does not belong);

an extractor for extracting a character string from the document, the character string having a length equal to the predetermined length of the plurality of text strings

contained in the database (Fig. 3 item S2-see above discussion of elements, C.7.lines 15-17); and

a language analyzer for determining a probability value fore each of the candidate languages that the character string does not belong to the candidate languages by retrieving the probability value associated to the character string from the database for each of the candidate languages, and includes logic for adjusting the negative assumption value based on the probability value, the language analyzer determining that the document is one language of the candidate languages when the negative assumption value passes a threshold value (C.7.lines 60-65, C.13.lines 44-58).

As per **claim 4**, Elworthy teaches claim 1, and further teaches further including an information retrieval engine for retrieving documents in response to a search request, the documents retrieved being analyzed by the language analyzer (C.13.lines 44-58).

As per **claim 5**, Elworthy teaches claim 1, and further teaches wherein the logic for adjusting includes logic for combining the negative assumption value (C.13.lines 44-58-his accumulated values) with the probability value (ibid, Figs 14a-c).

As per **claim 6**, Elworthy teaches claim 1, and further teaches wherein the language analyzer further includes iteration logic for causing the extractor to extract another character string from the document if the negative assumption value fails to pass the threshold value (C.12.lines 20-38).

As per **claims 7, 8 and 13**, Elworthy teaches a method of determining a language of a document from a set of candidate languages, the method comprising the steps of:

setting a null hypothesis to a true value for each candidate language indicating the document is not in the candidate language and setting a false value (C.12.lines 20-38-his setting of an initial confidence statistic and decision flag, claim 13);

extracting a text string from the document the text string having a predetermined length (see claim 1);

determining a contrary probability for each candidate language that the text string does not belong to the candidate language (see claim 1 as contrary probability value is interpreted as a negative assumption value) based on probabilities that the text string belongs to each of the candidate languages where the probabilities are retrieved from a database that stores probability values for a plurality of text strings each having the predetermined length, each text string of the plurality of text strings having an associated probability value for each candidate language indicating a probability that the text string occurs within a language from the candidate languages based on occurrences of the text string in all of the candidate languages (see claim 1);

adjusting the null hypothesis for each candidate language with the contrary probability corresponding to the candidate language (C.12.lines 20-38-his added value stored in the accumulator-as the null hypothesis); and

determining the document is one language from the candidate languages when the null hypothesis for the one language is disproved by approaching the false value

(C.12.lines 20-38, C.13.lines 22-35-the highest accumulated probability-accounts for approval and simultaneously disapproval, C.13.lines 44-58).

As per **claim 9**, Elworthy teaches claim 8 and further teaches repeating the extracting step for a different text string from the document and repeating the method until the null hypothesis is disproved for one of the candidate languages by passing the threshold value (C.12.lines 20-38).

As per **claim 10**, Elworthy teaches pregenerating probability data corresponding to each candidate language (C.2.lines 30-35-his "classification" as the candidate language), the probability data including a probability value for a text string that is normalized based on an occurrence probability of the text string in all the candidate languages (ibid, his "determined probability that an element or group of elements belongs to a classification" is interpreted as occurrence probability, C.2.lines 30-38, the comparison with probability values interpreted as the normalization).

As per **claim 11**, Elworthy teaches claim 7 and further teaches identifying the document based on a search request (C.1.lines 6-8-identifying a classification inherent to a search request).

As per **claim 12**, Elworthy teaches claim 7 and further teaches extracting a plurality of sequential characters that form the text string (C.2.lines 27-30, C.5.lines 6, 7).

6. Claims 15, 16, 20, and 21 are rejected under 35 U.S.C. 102(b) as being anticipated by Pon et al. (Pon, US 6,047,251).

As per **claims 15, 16 and 21**, Pon teaches a method of determining a language of a document from a set of candidate languages, the method comprising the steps of:

setting a probability assumption indicating the document is not in the candidate language(C.7.lines 36, 37-his setting of an initial "confidence statistic", C.7.lines 1, 2-his 1, as true, and 0, as false, value, claim 13);

extracting a text string from the document (C.7.lines 38-40);

disproving the probability assumption based on a contrary probability that the character string does not belong to the selected language (C.6.line 65-C.7.line 22-the "statistic that indicates whether a selected word is in a chosen language", wherein the "probability that a character string belongs to each of the candidate languages result inherently determines the value that a character string does not belong) such that if the contrary probability fails to support the probability assumption, then the document is determined as being in the selected language (C.7.lines 40-45-contrary and probability assumption, C.8.lines 1-4, his "region" as the document, his current subzone for the region "is likely to be the language of the region, C.8.lines 5-25-use of the threshold, C.9.lines 10-12-entire document, wherein the accumulation), and further determining the document is the selected language from a set of candidate languages (ibid, C.9.lines 14-32, claim 16).

As per **claim 20**, claim 20 sets forth limitations similar to claims 4 and 11, and therefore is rejected for the same reasons and under the same rationale.



***Claim Rejections - 35 USC § 103***

7. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

8. Claims 17-19 are rejected under 35 U.S.C. 103(a) as being unpatentable over Pon in view of Elworthy.

As per **claim 17**, Pon teaches claim 16, but lacks including generating a probability database having a contrary probability for each of a plurality of character strings for each of the candidate languages, where the contrary probability of a character string in one language is determined based on an occurrence frequency of the character strings in the one language influenced by a total occurrence frequency of the character string in all the candidate languages.

However, Elworthy further teaches generating a probability database having a contrary probability for each of a plurality of character strings for each of the candidate languages, where the contrary probability of a character string in one language is determined based on an occurrence frequency of the character strings in the one language influenced by a total occurrence frequency of the character string in all the candidate languages (C.8.lines 27-31-his "tokens" as character strings, Fig. 14a, b, c, C.13.lines 43-58-wherein the "probability" values inherently contain contrary probability values). Therefore, at the time of the invention, it would have been obvious to modify Pon's dictionary with Elworthy's lexicon/library (database) which contains probabilities

Art Unit: 2654

for each character string... The motivation for doing so would have been to identify a language/classification using predetermined values (abstract), and to develop a increasing accurate method in classifying data (C.2.lines 16-20).

As per **claim 18**, Pon and Elworthy make obvious claim 17, Elworthy further teaches determining the occurrence frequency of each character string based on a sample set of documents provided for each of the candidate languages (C.7.line 65-C.8.line 7).

As per **claim 19**, Pon and Elworthy make obvious claim 17, Elworthy further teaches wherein the contrary probability of the character string in one language is normalized by the total occurrence frequency of the character string in all the candidate languages (C.8.lines 27-31, C10.line 15-C.11.line 37, especially C.10.lines 50-57-his "frequency of all word tokens in M, and  $p(m)$  as the normalization).

As per **claim 14**, Claim 14 sets forth limitations similar to claims 17, 18, and 19, and is thus rejected for the same reasons, and under the same rationale.

### ***Conclusion***

9. Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire **THREE MONTHS** from the mailing date of this action. In the event a first reply is filed within **TWO MONTHS** of the mailing date of this final action and the advisory action is not

Art Unit: 2654

mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

10. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Lamont M. Spooner whose telephone number is 571/272-7613. The examiner can normally be reached on 8:00 AM - 5:00 PM.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil can be reached on 571/272-7602. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

  
**RICHEMOND DORVIL**  
**SUPERVISORY PATENT EXAMINER**

lms  
1/16/06